electronic reprint

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structures of *Micrococcus lysodeikticus* catalase, its ferryl intermediate (compound II) and NADPH complex

Garib N. Murshudov, Albina I. Grebenko, James A. Brannigan, Alfred A. Antson, Vladimir V. Barynin, Guy G. Dodson, Zbigniew Dauter, Keith S. Wilson and William R. Melik-Adamyan

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structures of *Micrococcus lysodeikticus* catalase, its ferryl intermediate (compound II) and NADPH complex

Garib N. Murshudov,[a]* Albina I. Grebenko,[b] James A. Brannigan,[a] Alfred A. Antson,[a] Vladimir V. Barynin,[b]† Guy G. Dodson,[a] Zbigniew Dauter,[c]‡ Keith S. Wilson[a] and William R. Melik-Adamyan[b]*

[a]Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, England, [b]Institute of Crystallography of Russian Academy of Sciences, Leninsky Prospect 59, 119333 Moscow, Russia, and [c]European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany

† Current address: The Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, First Court Western Bank, Sheffield S10 2TN, England.
‡ Current address: NCI, Synchrotron Radiation Research Section, Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA.

Correspondence e-mail: garib@ysbl.york.ac.uk, mawr@ns.crys.ras.ru

Received 2 May 2002
Accepted 22 July 2002

PDB References: MLC-0.88, 1gwe, r1gwesf; MLC-RT, 1hbz, r1hbzsf; MLC-PAA, 1gwf, r1gwfsf; MLC-NADPH, 1gwh, r1gwhsf.

The crystal structure of the bacterial catalase from *Micrococcus lysodeikticus* has been refined using the gene-derived sequence both at 0.88 Å resolution using data recorded at 110 K and at 1.5 Å resolution with room-temperature data. The atomic resolution structure has been refined with individual anisotropic atomic thermal parameters. This has revealed the geometry of the haem and surrounding protein, including many of the H atoms, with unprecedented accuracy and has characterized functionally important hydrogen-bond interactions in the active site. The positions of the H atoms are consistent with the enzymatic mechanism previously suggested for beef liver catalase. The structure reveals that a 25 Å long channel leading to the haem is filled by partially occupied water molecules, suggesting an inherent facile access to the active site. In addition, the structures of the ferryl intermediate of the catalase, the so-called compound II, at 1.96 Å resolution and the catalase complex with NADPH at 1.83 Å resolution have been determined. Comparison of compound II and the resting state of the enzyme shows that the binding of the O atom to the iron (bond length 1.87 Å) is associated with increased haem bending and is accompanied by a distal movement of the iron and the side chain of the proximal tyrosine. Finally, the structure of the NADPH complex shows that the cofactor is bound to the molecule in an equivalent position to that found in beef liver catalase, but that only the adenine part of NADPH is visible in the present structure.

1. Introduction

Catalase (hydrogen peroxide:hydrogen peroxide oxidoreductase; EC 1.11.1.6) is a redox enzyme present in all aerobic organisms (Deisseroth & Dounce, 1970). The enzyme's role is believed to be the protection of the cell from the toxic effects of hydrogen peroxide by decomposing it into water and molecular oxygen before highly reactive hydroxyl radicals are formed. Catalase deficiency is known to cause the inherited disease acatalasemia (Ogata, 1991). Catalases have been shown to be involved in alcohol oxidation (Bradford *et al.*, 1993) and play a protective role in inflammation (Halliwell & Gutteridge, 1984), in the prevention of mutations (Vuillaume, 1987) and in ageing and certain cancers (Mallery *et al.*, 1995). Catalase has been a subject of interest from the beginning of the last century and a number of recent reviews dealing with its biochemistry and structure are available (Bravo *et al.*, 1997; Zamocky & Koller, 1999; Nicholls *et al.*, 2001; Mate *et al.*, 2001).

Three groups of catalases have been identified. The first, most fully characterized, group contains the mono-functional

'classic' haem-containing catalases, which are homotetramers with one haem per subunit and are present in both prokaryotes and eukaryotes. The subunit sizes within this group vary, consisting of about 500 residues (small-subunit catalases) or 700 residues (large-subunit catalases). The second group is composed of the non-haem Mn-catalases, homohexameric enzymes with two manganese ions in the active site, that have so far been found only in prokaryotes (Antonyuk *et al.*, 2000; Barynin *et al.*, 2001). The third group, haem-containing catalase–peroxidases with dual catalytic activity, are apparently gene-duplicated members of the plant peroxidase superfamily (Welinder, 1991).

The haem-catalases are not only exceptional mediators of hydrogen peroxide decomposition but are also effective catalysts of peroxide-dependent oxidation of hydrazoic, formic and nitrous acids, of lower aliphatic alcohols and of hydroxylamine (Schonbaum & Chance, 1976). The reactions occur in at least two steps:

$$(1)\quad E(\text{porph-Fe}^{III}) + R\text{OOH} \rightarrow \underset{\text{Compound I}}{E(\text{porph}^{+}\text{-Fe}^{IV})=O} + R\text{OH}$$

$$(2)\quad \underset{\text{Compound I}}{E(\text{porph}^{+}\text{-Fe}^{IV})=O} + H\text{X}OH \rightarrow E(\text{porph-Fe}^{III}) + H_2O + \text{X}O$$

or

$$\text{Compound I} + e^{-} \rightarrow \text{Compound II}$$

Where E is the enzyme, porph is protoporphyrin IX, *R* is H, alkyl or acetyl and *X* is O, C═O or $(CH_2)_n$. When *R* is H and *X* is O, ROOH and H*X*OH become the best known substrate of catalase: H_2O_2.

The first step of the reaction is a two-electron oxidation of the resting ferric state of the enzyme with an Fe^{III}-porphyrin group to compound I, which contains an Fe^{IV}-oxo porphyrin with a π-cationic radical (Dolphin *et al.*, 1971; Roberts *et al.*, 1981). The second step is a two-electron reduction of compound I back to the resting state directly or *via* an alternative pathway through compound II, produced by a single-electron reduction of compound I. Compound II may then return to the ferric state using a one-electron donor or may be converted to inactive compound III by a further two-electron oxidation by H_2O_2. Some catalases are more resistant to H_2O_2-induced inactivation *via* formation of compound II and compound III than others (DeLuca *et al.*, 1995; Lardinois *et al.*, 1996). For example, catalases from *Penicillium vitale* (PVC), *Escherichia coli* (hydrogenperoxidase II; HPII) and the large-subunit catalases do not form compound II at all. Most of those that are able to form compound II are small-subunit catalases that apparently bind NADPH. This prevents compound II formation *via* one- or two-electron transfer from NADPH to the haem (Almarsson *et al.*, 1993; Kirkman *et al.*, 1999; Hillar *et al.*, 1994; Olson & Bruice, 1995).

M. lysodeikticus catalase (MLC) is one of the 'classic' haem-containing catalases. At present, the crystal structures of seven catalases of this group are known from different species and at different resolutions: those from beef liver (BLC; Fita *et al.*, 1986), *P. vitale* (PVC; Vainshtein *et al.*, 1986; Murshudov *et al.*, 1996), *M. lysodeikticus* (MLC; Murshudov *et al.*, 1992), *E. coli* (HPII; Bravo *et al.*, 1995), *Proteus mirabilus* (PMC; Gouet *et al.*, 1995), *Saccharomyces cerevisiae* (SCC-A; Mate *et al.*, 1999) and human erythrocyte (HEC; Putnam *et al.*, 2000).

The structure of MLC was reported previously at 1.5 Å resolution with the X-ray amino-acid sequence derived from the electron-density map (Murshudov *et al.*, 1992). Here, we present the revised amino-acid sequence of MLC deduced from the cloned gene, together with four crystal structures: the room-temperature structure at 1.5 Å resolution (MLC-RT), the atomic resolution cryogenic structure at 0.88 Å (MLC-0.88), the cryogenic structure of MLC compound II at 1.96 Å (MLC–PAA), and the room-temperature structure of MLC complexed with NADPH at 1.8 Å (MLC–NADPH).

2. Materials and methods

2.1. DNA cloning and sequencing

For direct cloning and expression, the degenerate oligonucleotides 5′-d(TTTT**CCATGG**AGCACCAGAAGAC*S*AC-*S*CC*S*CAC)-3′ and 5′-d(TTTT**GAATTC**CACAT*S*CG*S*GCCT-C*S*CC*S*CC*S*GCCTC)-3′ (where *S* = C or G) were synthesized (0.2 μ*M* scale) using an ABi392 machine. These were designed to match the N-terminal chemically derived sequence (TEHQKTTPHA) starting at Glu2 and the X-ray derived C-terminal sequence, respectively, although it should be noted that the refined native MLC structure reveals an Ile residue at the C-terminus, not Met as previously reported. PCR was performed in 50 m*M* KCl, 10 m*M* Tris–HCl pH 8.3 buffer using 10% glycerol co-solvent and final concentrations of 200 μ*M* dNTP, 3 μ*M* $MgCl_2$, 1 μ*M* degenerate primer. 25 cycles of 369 K for 90 s and 345 K for 150 s were performed using 200 ng of *M. luteus* ATCC4698 chromosomal DNA as template and 2.5 U Taq polymerase, yielding a single product of 1.5 kbp flanked by *N*coI and *Eco*RI restriction sites (bold) for cloning into pET21d (Novagen) of the T7 expression system of Studier (1991). Single-stranded DNA derivatives of the recombinant pET clones were prepared by transfection with helper phage R408 (Promega) to obtain the transcribed strand or by capture of biotinylated PCR products on magnetic beads (Dynal) for the coding strand. DNA was sequenced using a battery of oligonucleotides which prime at regular intervals, extended in the presence of labelled dCTP to enhance signal strength by T7 polymerase with deazaG/A mixes (Pharmacia) and at elevated temperature (338 K) using a thermostable enzyme (Iso-therm, Epicenter Technologies) to relieve compressions. The sequence was confirmed on one strand using dye-terminators and thermal cycling incorporating dITP on an automated sequencer (ABi 373). A number of CT transversion mutations were evident, consistent with the known mutational tendency of Taq polymerase, and to which the extended denaturation phase during the PCR may have been a contributing factor. A composite DNA sequence based on independent isolates has been deposited at the EMBL under accession code AJ438208.

2.2. Purification and crystallization

The purification and crystallization of *M. lysodeikticus* catalase have been described previously (Herbert & Pinsent, 1948; Marie *et al.*, 1979). In brief, native crystals were grown by hanging-drop vapour diffusion from 15–18 µl droplets containing 20–30 mg ml^{-1} protein, 0.6–0.8 *M* ammonium sulfate and 0.05 *M* sodium acetate buffer pH 5.2 over a reservoir precipitant solution containing 1.2–1.4 *M* ammonium sulfate in the same buffer (Murshudov *et al.*, 1992). Crystals suitable for X-ray analysis, with dimensions of 0.4–0.6 mm, grew after one week. The crystals belong to the space group $P4_22_12$, with unit-cell parameters a = 106.7, c = 106.3 Å. Before vitrification, crystals of MLC were soaked for about 5 s in a solution of 25% glycerol, 1.4 *M* ammonium sulfate and 0.1 *M* sodium acetate buffer pH 5.2.

The formation of compound I and compound II of MLC was followed spectrophotometrically in a solution containing 0.1 mg ml^{-1} MLC, 0.1 m*M* peracetic acid (PAA), 1.2 *M* ammonium sulfate and 0.05 *M* sodium acetate buffer pH 5.2. Compound II formation was complete after 2 min and it was stable for 24 h.

Compound II crystals were prepared by soaking native crystals in mother liquor with addition of 125 m*M* PAA and observing their change in colour. Immediately after soaking, the crystal changed from brown to green, indicating that formation of compound I had begun. After 3 min, formation of compound II was seen to be complete as the colour changed to red, whereupon the crystal was vitrified at 120 K.

Spectral changes in solutions of MLC caused by PAA in the presence and absence of NADPH show that it prevents formation of compound II as for BLC and other NADPH-containing catalases. Crystals of MLC complexed with NADPH (MLC–NADPH) were grown under similar conditions to those of native enzyme with the addition of NADPH to the protein droplet to a concentration of 4 m*M*.

Table 1
Statistics of data collection and structure refinement.

Crystal	MLC-0.88†	MLC-RT‡	MLC–PAA§	MLC–NADPH¶
Space group	$P4_22_12$	$P4_22_12$	$P4_22_12$	$P4_22_12$
Unit-cell parameters (Å)	105.8, 105.0	106.7, 106.25	106.19, 105.3	106.7, 106.25
Data-collection statistics				
No. unique reflections	457407	95355	43789	54476
Resolution (Å)	74.5–0.88	12–1.5	20–1.96	37–1.75 (1.83)††
Outer resolution shell (Å)	0.897–0.88	1.54–1.50	1.98–1.96	1.80–1.75
R_{merge} (overall) (%)	5.1	6.1	8.0	7.3
R_{merge} (outer shell) (%)	24.9	20.8	38.9	9.5
Multiplicity (overall)	3.5	3.7	5.5	5.7
$\langle I\rangle/\langle\sigma\rangle$ (overall)	24.4	8.0	9.0	16.0
$\langle I\rangle/\langle\sigma\rangle$ (outer shell)	1.3	3.5	2.0	4.0
Completeness (overall) (%)	99.3	97.5	99.0	87.7
Completeness (outer shell) (%)	88.8	96.4	99.0	43.7
B_{Wilson}	6.0	12.0	22.7	15.0
Statistics of refined models‡‡				
R value (%)/No. reflections	8.9/433749	9.5/90574	14.1/41589	11.1/51736
R_{free} (%)/No. reflections	9.6/22913	12.1/4781	19.0/2200	13.8/2740
DPI values§§ (Å)	0.009	0.046	0.129	0.078
DPI_{free} values§§ (Å)	0.009	0.042	0.128	0.077
G factor	0.06	−0.11	−0.23	−0.21
No. protein atoms	4213	4042	4052	4033
No. solvent atoms	878	538	711	517
No. NADPH	0	0	0	1 (0.5)
No. sulfate ions	1 + 2 × 0.5	1	3	1
No. acetate ions	0	0.5	2 × 0.5	2 × 0.5
R.m.s.d. of bond lengths (Å)	0.016	0.013	0.022	0.014
R.m.s.d. of bond angles (°)	1.95	1.5	1.8	1.6
R.m.s.d of planar angles (Å)	6.6	6.0	6.5	6.0
$\langle B\rangle$ main chain (Å^2)	1.3	9.4	22.3	12.4
$\langle B\rangle$ side chain (Å^2)	2.5	13.1	26.1	17.3
$\langle B\rangle$ waters (Å^2)	18.1	27.8	41.0	33.6

† MLC-0.88, native data collected at 110 K. ‡ MLC-RT, native data collected at room temperature. § MLC–PAA, data from crystals soaked in PAA solution and frozen to 120 K. ¶ MLC–NADPH, data collected at room temperature from MLC co-crystallized with NADPH. †† Effective resolution is shown in parentheses. Data were collected using a Rigaku R-AXIS detector, which has a rectangular shape. All data were processed and used in the refinement. As a result, the outer shell has lower completeness. The effective resolution (Cruickshank, 1996) was calculated using the equation $d_{min}C^{-1/3}$, where d_{min} is the highest resolution and C is the completeness of the data. ‡‡ In the refinement, ideal bond lengths and angles (Engh & Huber, 1991) were used as target values. §§ DPI (diffraction data-precision indicator) is an estimate of coordinate error (Cruickshank, 1996).

2.3. Data collection

The first data set from the native (or resting-state) enzyme at 1.5 Å resolution (MLC-RT) was collected at room temperature at a wavelength of 0.91 Å using synchrotron radiation at the EMBL Hamburg Outstation and a MAR Research image-plate detector on beamline X31. The data were processed with the *MOSFLM* program package (Leslie, 1992) and merged using the *ROTAVATA*/*AGROVATA* programs (Collaborative Computational Project, Number 4, 1994).

High-resolution data from the native enzyme to 0.88 Å (MLC-0.88) were collected using synchrotron radiation at the BW7B beamline (EMBL, Hamburg) at a wavelength of 0.89 Å from a frozen crystal at 110 K. X-ray data from a crystal of compound II complex at 1.96 Å (MLC–PAA) were collected at 120 K using Cu $K\alpha$ radiation from a Rigaku RU-200 rotating-anode generator and MAR Research image plate. The same experimental setup was used to collect data from the dinucleotide complex of MLC with NADPH, except that an R-AXIS IIC image-plate detector was used. The latter data set was collected to 1.83 Å resolution at room temperature. All three data sets were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Statistics of the data sets are given in Table 1.

2.4. Refinement

With the exception of MLC-RT, where initial calculations were performed using *PROLSQ*, all refinements were performed using maximum-likelihood procedures as imple-

mented in the program *REFMAC* (Murshudov *et al.*, 1997) available from *CCP*4 (Collaborative Computational Project, Number 4, 1994). In the later stages, all waters were checked using the X-fit option of *QUANTA* (Accelrys Inc.) and new water molecules were added using the X-solvate option (Oldfield, 1996). In the final stages, H atoms were added at their riding positions. All measured reflections except those chosen for $R_{\rm free}$ and σ_A calculation were used. For all four structures, overall anisotropic scale factors were refined and applied to F_c. Where multiple conformations were suggested by the difference map, the B values of the relevant atoms were checked to see if they provided consistent information. If so, a second, alternative, conformation was introduced.

2.4.1. MLC-RT. The initial refinement of MLC-RT using *PROLSQ* (Hendrickson & Konnert, 1980) gave an R value of 15.0% (Murshudov *et al.*, 1992). In the first stages, waters were added using the program *WATPEAK* (Vagin *et al.*, 1998) and *ARP* (Lamzin & Wilson, 1993). For further refinement, 5% of the data were chosen for free R value (Brünger, 1993) validation. The model was rebuilt using the amino-acid sequence deduced from the gene. Up to this stage, all reflections in the resolution range 8–1.5 Å except those chosen for free R-value calculation were used for refinement and for map calculation. At this point, maximum-likelihood refinement implemented in *REFMAC* became available and all subsequent refinement was carried out using this program and using all measured reflections. The resulting MLC-RT model was used as a starting point for refinement of the other three structures.

2.4.2. MLC-0.88. After ten cycles of isotropic refinement, $R/R_{\rm free}$ reached 16.5/17.2%. Further refinement was carried out using individual anisotropic atomic refinement in *REFMAC* (Murshudov *et al.*, 1999). After every ten cycles of refinement, all protein atoms were checked to look for alternative conformations. All water atoms were inspected using $F_o - F_c$ maps. If there was negative density at the position of a water atom then the $2F_o - F_c$ map was inspected to see if there was atom at this position. If there was clear density in the $2F_o - F_c$ map then its occupancy was reduced to 0.5. Two additional sulfate ions were located with partial occupancy.

2.4.3. MLC-PAA. After the first cycles of refinement against the cryogenic MLC–PAA data, the difference map showed clear density at the sixth coordination position of the iron and this was introduced to the model as an O atom. The relatively poor data and refinement statistics compared with the native crystals (Table 1) are probably the result of deterioration of the crystal, perhaps a consequence of the long soaking time in the PAA solution.

2.4.4. MLC-NADPH. After the first ten cycles of refinement, $F_o - F_c$ and $2F_o - F_c$ maps showed clear density for the adenine base, one ribose and two phosphates of the NADPH. The nucleotide was built into the density using *QUANTA*.

2.4.5. H atoms. In all four structures, the addition of H atoms in their riding positions reduced the $R/R_{\rm free}$ (in MLC-RT by 1.4/1%, in MLC–NADPH by 1.4/0.8% and in MLC–PAA by 0.8/0.2%). For MLC-0.88, H atoms were included from the start. The fall in $R_{\rm free}$ is greater at high resolution, indicating that the protein atomic positions are accurate.

Unless otherwise stated all crystallographic calculations were carried out using programs from the *CCP*4 program suite. All refined models were checked with *PROCHECK* (Laskowski *et al.*, 1993). Final statistics of the refinement are given in Table 1.

3. Results and discussion

3.1. Gene cloning and primary structure

The PCR approach was used for the direct cloning of the catalase gene from chromosomal *M. luteus* DNA (this strain is synonymous with *M. lysodeikticus*). The method is an extension of protocols that utilize degenerate primers based on conserved amino-acid sequences in multiple alignments of coding regions to prime the PCR. In this case, the primers were designed to both amplify a complete gene, based on the chemically determined N-terminal sequence and the C-terminal residues visible in the three-dimensional structure (Murshudov *et al.*, 1992), and to incorporate unique restriction sites for cloning (see §2). The degeneracy of the primers was minimal, as the choice of synonymous codons are biased towards those which maximize their G + C content, reflecting the low A + T content of the *M. luteus* genome (∼30%). For example, of the six possible codons that specify leucine, 99% are CTG or CTC. Although it is unusual to know both the N- and C-terminal protein sequence for a gene-cloning target, advances in C-terminal protein sequencing suggests that such an approach will become more general. The recombinant catalase protein overexpressed in *E. coli* runs at the equivalent position to authentic *M. lysodeikticus* catalase on SDS–PAGE (data not shown), although the specific activity is lower than expected, presumably owing to the restricted availability of haem cofactor. The cloning of the catalase gene opens up the possibility of using site-directed mutagenesis of this protein for structure–function analysis. The revised MLC protein sequence as translated from the catalase gene is 84% identical to that derived from the previous X-ray model. The differences are minor (*e.g.* Val *versus* Thr) and are often at positions located on the surface of the molecule. However, some alterations to the X-ray-derived sequence proved crucial for correct model building owing to ambiguities caused by alternate conformations (*e.g.* residues Leu105 and Tyr378, see below).

3.2. Analysis of MLC structures

3.2.1. Overall fold. The overall structure of MLC has been described previously (Murshudov *et al.*, 1992). It has a three-dimensional organization that closely resembles that of BLC and other known catalases. The DNA-deduced amino-acid sequence of MLC shows that the previous numbers of residues assigned by X-ray analysis should be increased by four.

MLC is a homotetramer with 222 molecular symmetry (Fig. 1*a*). Each monomer consists of 503 residues and one protoporphyrin IX group. Like BLC (Fita *et al.*, 1986), MLC has four domains. The first N-terminal domain (1–58) contains one α-helix, which is completely surrounded by residues of

neighbouring subunits within the tetramer. The second domain (59–352) is an $\alpha+\beta$ domain which contains eight α-helices and an eight-stranded β-barrel. This domain and the fourth four-helical domain (425–503) are connected by the third domain (353–424) which wraps around the surface of the molecule and together with the first domain is responsible for the organization of the tightly packed tetramer.

A comparison of the MLC and BLC folds is shown in Fig. 1(*b*). Their three-dimensional similarity starts at residue 13 of MLC and residue 26 of BLC and ends at residues 487 and 499, respectively. The r.m.s. deviation between CA atoms of the 463 equivalent residues is 1.1 Å, as calculated using *O* (Jones *et al.*, 1991) for residue pairs where the difference between CA atoms was less than 3.8 Å. 210 of these residues are identical in sequence. The biggest differences in the CA position, apart from the N-terminus, are in the wrapping domain, at the 401st (8 Å) and at the 422nd (9 Å) residues of MLC (Fig. 1*b*).

3.2.2. Resting state: MLC-RT and MLC-0.88. As the statistics suggest, the MLC-0.88 data (Table 1) are of very high quality. Most of the H atoms could be identified directly in the $F_o - F_c$ map at the 3σ level. The most important of them from a mechanistic point of view – those of the active site His61 and Asn133 – are discussed below. In spite of the high quality of the data and exhaustive refinement, positive $F_o - F_c$ electron-density features remain around the molecule and in the channels associated with water molecules.

As the conformational angles φ and ψ were not restrained during refinement, they can act as independent indicators of model quality. The conformational angles of 99.6% residues are within the allowed regions of the Ramachandran plot (Fig. 2; Ramakrishnan & Ramachandran, 1965). The φ and ψ angles of only two residues, Val22 and Ser202, are in the disallowed region. Both are the central residues of γ turns. Val22 takes part in the intersubunit interaction and despite its hydrophobicity is unusually exposed: the accessibility surface area for CG1 is 11 Å^2 and for CG2 is 34 Å^2. The unusual conformation of Ser202 is a common feature in all known catalase structures and seems to be related to the folding of the barrel.

Four other non-glycine residues have positive φ and ψ angles. One of them, His61, is conserved in the other catalases and is essential for catalytic activity. The other three are situated either in the connecting segment (Met376 and Glu373) or in the C-terminal region (His487).

Out of 26 prolines, two are in the *cis* conformation. One (Pro60) is situated before the active-site His61 and seems to be a special feature of MLC. Preceding the active-site histidine, most catalases contain the sequence Val-Val. In MLC these residues are Arg-Pro. Pro60 is probably in the *cis* conformation to avoid unfavourable interactions between the long hydrophilic side chain of Arg59 and the hydrophobic part of the haem. The second *cis* proline, Pro389, is a feature of most catalases (Gouet *et al.*, 1995; Fita *et al.*, 1986; Bravo *et al.*, 1995) and is believed to play a role in the tetramer organization, as there are a number of intersubunit interactions around this residue.

No electron density could be identified for the first five N-terminal residues. There are negative difference electron densities around residues 6–11. At this resolution, *B* values in general do not compensate for inaccurate occupancies. It is clear that these residues have several conformations, of which only one has been successfully modelled. MLC-0.88 shows that going towards the N-terminus along the polypeptide chain, disorder starts from residue 11 and becomes most pronounced at residue 6. The rest of the structure is mostly well defined with low *B* values.

The comparison of room-temperature and low-temperature structures of MLC

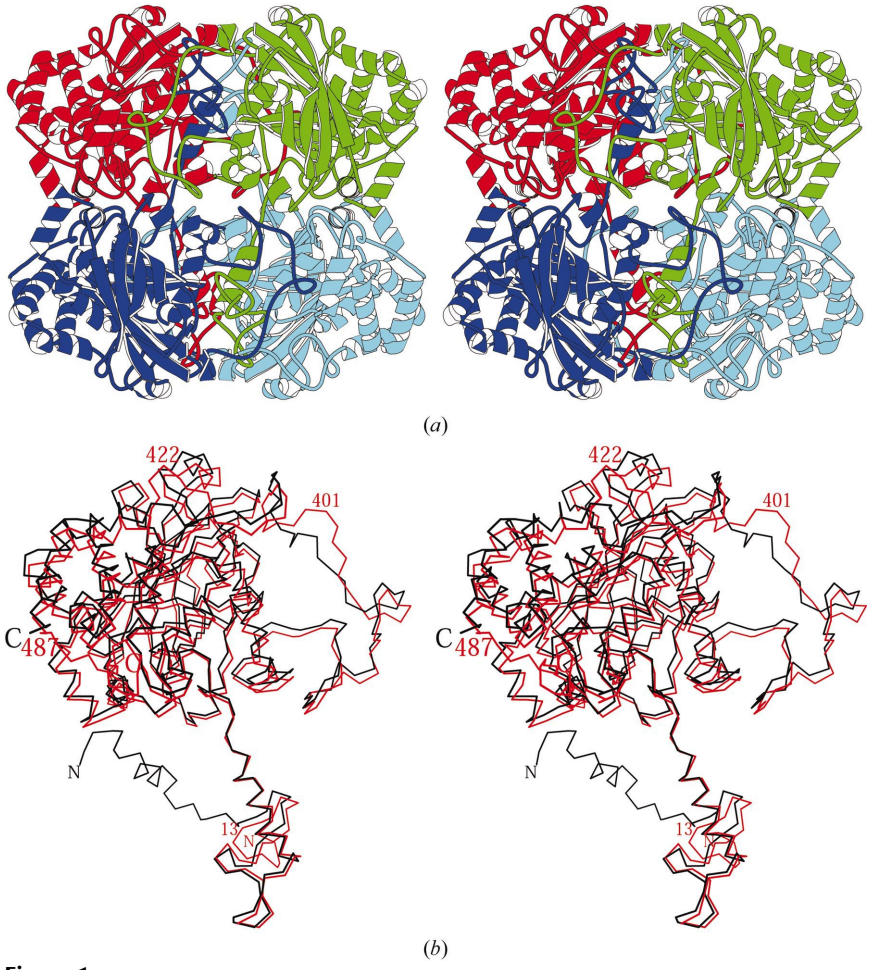

Figure 1
Stereoview of (*a*) ribbon diagram of the MLC tetramer and (*b*) CA trace of the *M. lysodeikticus* catalase monomer (in red) superimposed with that of beef liver catalase (in black). Figures were generated using *MOLSCRIPT* (Kraulis, 1991).

(MLC-RT and MLC-0.88) shows that there are no significant differences between them, apart from the number of waters and double conformations visible. The r.m.s. deviation of CA atoms is 0.185 Å between the two structures.

The haem geometry is the same in the two structures, showing similar bending, and the Fe atom is displaced from the NA–NB–NC–ND plane towards the proximal side by a similar distance. As expected, more waters (878) are detected in the ultrahigh-resolution MLC-0.88 structure compared with the MLC-RT structure (538). This partly reflects the high resolution and quality of the MLC-0.88 data and partly reflects the fact that local minima become more pronounced in vitrified crystals. Of special interest are waters in the active site and in the haem channel in MLC-0.88 (see below). In addition to one sulfate ion bound in MLC-RT, MLC-0.88 has two additional sulfate ions, each half occupied.

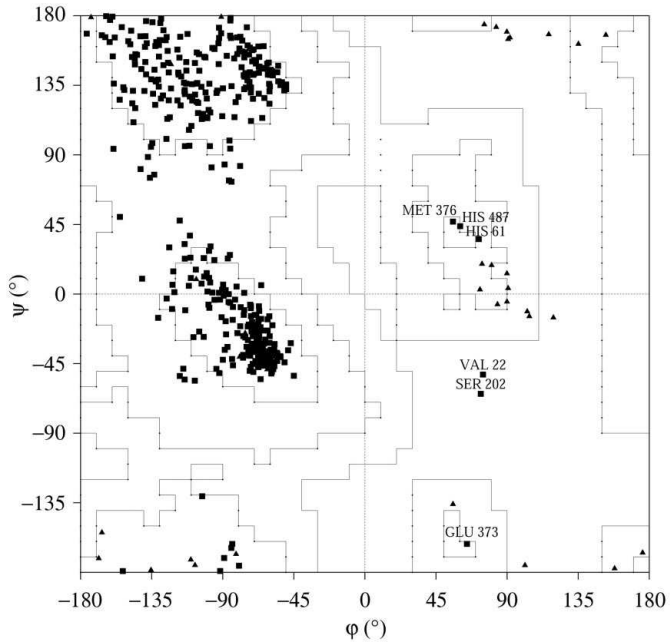

Figure 2
Ramachandran plot (Ramakrishnan & Ramachandran, 1965) for the *M. lysodeikticus* catalase model. Glycine residues are indicated by triangles and all other residues by squares. Residues with positive φ angles are highlighted. 88.5% of residues are in the most favoured regions and 11.1% are in allowed regions. Two residues, Val22 and Ser202, are in the disallowed region.

For similar reasons, there are more residues with double conformations detected in MLC-0.88 (45) than in MLC-RT (12). Only eight are common to both structures. There are two regions with multiple conformations, which are of especial interest as they break down the exact 222 molecular and crystallographic symmetry and introduce minor heterogeneity into the homotetramer. The first is the main chain of residues 105–106 (Fig. 3). The density clearly shows double conformations for these atoms and correlated double conformations for the side chain of Leu105. The two conformations involve a 180° flip about the peptide bond. The two conformations are complementary in the two subunits. As seen in Fig. 3, the conformation with the carbonyl O atom pointing towards the opposing subunit cannot simultaneously coexist in the symmetry-related subunit without giving rise to serious steric conflicts. These residues (105–106) are adjacent to Ser99, which is conserved in all catalases. Ser99 OG makes a strong hydrogen bond (2.78 Å) with the essential His61 and is presumably responsible for stabilizing its orientation. This feature (105–106) of double conformation is detected in most catalases refined with high-resolution data (*e.g.* PVC and HEC) and it is probable that it is of functional importance, particularly as the residues Val101 and Ala102 lie along the edge of the haem channel.

The second residue that violates the quaternary 222 symmetry of the tetramer is Tyr378. There are again two complementary conformations of the side chain in adjacent subunits, with the same conformation in each being in serious steric conflict. Tyr378 is distant from the active site and is not conserved among catalases. It is a feature specific to MLC and almost certainly plays no significant role in catalytic activity.

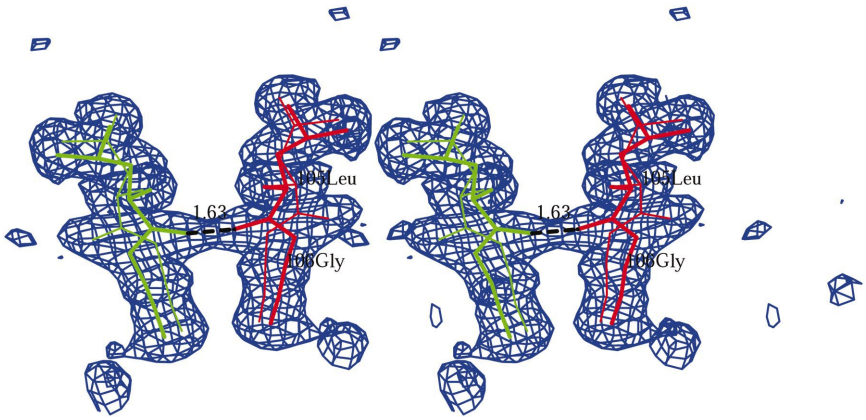

Figure 3
Double conformations of residues Leu105 and Gly106 that violate both crystallographic and molecular symmetry. Atoms from different subunits are coloured red and green. Reference conformations are shown in bold lines, with the alternative conformations in thin lines. Electron density was calculated using MLC-RT data and contoured at a level of 1σ.

3.2.3. Compound II. The data for the compound II complex extend to 1.96 Å and its analysis is correspondingly less detailed. There are no major changes in the overall structure relative to the resting-state molecule and the r.m.s. deviation between CA atoms of MLC–PAA and MLC-0.88 is 0.16 Å, and those of MLC–PAA and MLC-RT is 0.15 Å. The largest differences are in the N-terminal residues 6–7, which have high *B* values, and in the surface loops (residues 250–260 and 373–390), where there is some disorder. Five out of seven methionines have been oxidized to form methionine sulfoxide, but it is unlikely that this plays any functional role. This oxidation probably arises from prolonged soaking of the crystals in a solution of high PAA

Table 2
Changes in the active site of MLC structures.

Crystal	C3B—CHC—C2C angle† (°)	C3D—CHA—C2A angle‡ (°)	W*§—FE distance (Å)	W*§FE *B* factors (Å^2)
MLC-0.88	165.33	169.03	2.31	5.83/3.54
MLC-RT	168.44	162.69	2.40	14.18/6.26
MLC–PAA	157.77	167.21	1.87	20.01/20.88
MLC–NADPH	171.33	168.05	2.19	15.06/9.31

† C3B—CHC—C2C is the angle between pyrroles *B* and *C*. ‡ C3D—CHA—C2A is angle between pyrroles *D* and *A*. § W* is water W289 for MLC-0.88, MLC-RT and MLC–NADPH and is the O atom bound to the haem iron for MLC–PAA.

concentration. One of the oxidized methionines (Met185) is close to the haem channel, three others (Met53, Met144 and Met376) are close to the residues from symmetry-related subunits and one (Met166) is close to the surface. Analysis of the molecular surface shows that all of these residues are accessible to water and/or substrates. Inspection of the electron density showed that there are some regular features near to them. One of these (near to Met185) could be described as an acetate ion, which is a byproduct of methionine oxidation by PAA.

There are fewer water molecules detected in MLC–PAA than in MLC-0.88, MLC-RT or MLC–NADPH, owing to the relatively low resolution. Like MLC-0.88, MLC–PAA has two extra half-occupied sulfate ions that are not seen in MLC-RT.

In MLC–PAA there are eight residues with multiple conformations, fewer than in MLC-RT and MLC-0.88; however, the symmetry-violating residues are disordered in all molecules.

His49 has two conformations in MLC-RT and MLC–NADPH and only one in MLC–PAA and MLC-0.88. In the MLC-RT structure there is a correlated water molecule, spread over two sites. In the PMC compound I structure at 2.7 Å resolution (Gouet *et al.*, 1996) there was some strong electron density near to His42, equivalent to His49 in MLC, which was interpreted as a superoxide ion. Analysis of MLC compound II $2F_o - F_c$ and $F_o - F_c$ maps showed no evidence for two conformations for His49 and the associated water, nor for a superoxide ion.

The most interesting changes in compound II occur at the haem iron and the adjacent water molecule, as discussed below.

3.2.4. Complex with NADPH. The MLC–NADPH structure at a resolution of 1.83 Å shows that the nucleotide binds at an equivalent site to that seen in BLC, but that only the ribose, base and phosphate of the adenine have well defined density. The phosphate of NADPH has replaced the sulfate ion present in the resting-state structures (Fig. 4). There is no density for the nicotinamide base, second ribose and third phosphate, but there is density near to a possible position for the third phosphate of NADPH. However, this could also be a sulfate ion, since sulfate was present in the crystallization solution.

The adenine base of the NADPH lies parallel to Arg188, which is conserved among all known NADPH-containing catalases, binding NADPH through its phosphate O atom OP1. In BLC and other catalases, the first phosphate group is further stabilized by hydrogen bonding to a positively charged residue (Lys236 in BLC, Lys237 in HEC and Arg216 in PMC). In MLC, Ile222, which avoids any close contact with a phosphate group, is situated at this equivalent position and the additional interaction is with MLC Arg295. In other NADPH-containing catalases Arg295 is replaced by the hydrophobic residue leucine, causing a shift in phosphate position, which in MLC has moved 2.23 Å from the equivalent BLC position and some 3 Å from that in PMC. In spite of this difference, the adenine base is situated in the same position in all the enzymes. The nicotinamide moiety of NADPH in the MLC complex may have a position different from that in BLC, possibly related to the replacement of His304 involved in nicotinamide binding in BLC by Gln290 in MLC. His304 is conserved among other NADPH-containing catalases and binds the nicotinamide base *via* a hydrogen bond to NO7. These differences may explain why the nicotinamide moiety of NADPH is not visible in the MLC complex.

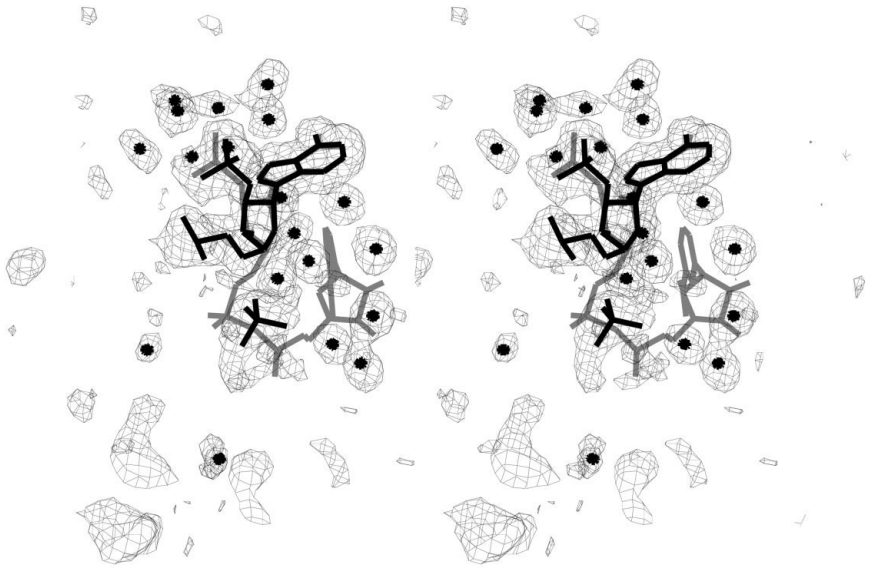

Figure 4
Electron density and model corresponding to NADPH. For comparison, the position of NADPH binding to beef liver catalase (grey) is also shown. Atoms of NADPH bound to *M. lysodeikticus* catalase are shown in black. Water molecules and the sulfate ion occupying the proposed NADPH-binding site are also shown.

Other parts of the structure are very similar to that of the resting state, with a r.m.s. deviation between CA atoms of MLC–NADPH and MLC-0.88 of 0.197 Å, and between MLC–NADPH and MLC-RT of 0.063 Å.

3.2.5. The active site in the resting state. The atomic resolution study has brought to light details of haem conformation, the hydrogen bonding and other interactions at the active site. The haem is distinctly bent

along a line that bisects the propionic acids in both molecules. The distances from the haem iron to the nearest water or O atom (in compound II) are presented in Table 2. In MLC-RT the water molecule has a relatively high *B* value. In MLC-0.88, only when this Fe-linked water was assigned partial occupancy does its *B* value become comparable to that of surrounding protein atoms. The Fe is displaced from the plane of the pyrrole N atoms by 0.14 Å towards the Tyr343 phenolate O atom. The close hydrogen bonds (2.7 Å) to the phenolate O atom by Arg339 NE and NH2 favour the loss of a proton through electrostatic influence and through steric effects. It is clear that the phenolate O atom is deprotonated both from the bond distance to the Fe and from the geometry of the oxygen's interactions with the N atoms of Arg339. There is an extensive hydrogen-bonding network extending from the phenolate O atom to Arg339, to His203 and to Asp333, and from here to the surface of the molecule. Inspection of the electron density shows the position of the H atoms in many of these interactions and confirms the description of the networks made in earlier reports (Fita & Rossmann, 1985).

The active site of the catalases is well characterized and is situated at the distal side of the haem. In MLC it includes the essential residues His61, Asn133 and Ser99. Unlike BLC (Fita *et al.*, 1986), the haem in MLC is well defined and free of any damage. His61 donates a proton in its hydrogen bond to Ser99 OG (adjacent to the symmetry-violating double conformation of residues 105–106), which is responsible for maintaining the orientation of the histidine. The active-site Asn133, unlike His61, does not make any strong hydrogen bonds and it is not clear what contributes to the stabilization of its orientation. The closest contacts that Asn133 makes are to the waters W36 in the active site (3.20 Å) and W50 (3.10 Å) hydrogen bonded with Ser202 with an unusual (φ, ψ) conformation. Experimentally observed H atoms on His61 and on Asn133 ND2 (Fig. 5) clearly confirm that the current orientations of these residues are correct. Fita & Rossmann (1985) based their reaction mechanism on this orientation of these residues using 2.5 Å resolution data where their orientation was less certain. Small positive density close to His61 NE2 suggests partial protonation of this residue (Fig. 5); this could have implications for the mechanism.

Although it was shown by NMR (Lanir & Schejter, 1975) that the sixth coordination position of the iron could be occupied by a water molecule, experimental results using magnetic circular dichroism suggest that this position is empty (Andersson *et al.*, 1995). In the MLC structure there is a partially occupied water W289 at a distance of 2.31 Å from the iron, occupying a near-ideal octahedral coordination position, although the bond is a little longer than expected (around 2 Å). No corresponding water molecule was reported in BLC (Fita *et al.*, 1986) or PMC (Gouet *et al.*, 1995), although there is a water molecule in PMC at a much larger distance from the iron, about 2.8 Å. However, the high-resolution structures of PVC (Murshudov *et al.*, 1997) and HPII (Bravo *et al.*, 1995) both contain a water at a corresponding position with an iron–water distance of about 2.28 Å. Inspection of the electron density for the HEC structure at 1.5 Å resolution shows a low-density water peak at about 2.3 Å from the iron, suggesting a partially occupied water as in the MLC structures. The disagreement between crystallographic and magnetic circular dichroism data might be explained by the fact that this water is a very weak ligand of the iron, which stays in a high-spin state. A second water, W36, with a *B* value of 9.1 Å^2, lies close by and makes a strong hydrogen bond with the iron's weakly liganded water (the W36—W289 distance is 2.57 Å) and with the essential His61 (distance 2.67 Å). There is a long and weak hydrogen bond between this water and ND2 of Asn133 at a distance of 3.19 Å. The high *B* value of W36 relative to neighbouring atoms (the average *B* value for this region is 5 Å^2) and the partial occupancy of W289 might suggest that during the course of reaction they are easily displaced by substrate.

Table 3
Comparison of angles (°) between pyrrole rings in the resting state of MLC-0.88 and in MLC–PAA.

	A	*B*	*C*	*D*
A	0.0	2.5/2.4	21/21	16/12
B	2.5/2.4	0.0	21/23	17/14
C	21/21	21/23	0.0	4/10
D	16/12	17/14	4/10	0.0

The partially occupied iron-coordinating water and the double conformation of residues 105–106 may correlate with the observation that only half of the catalase is active at the same time (Chance & Oshino, 1971), although the structure of compound II shows that all Fe atoms are bound to O.

3.2.6. Comparison of the active site in compound II and the resting state. The changes associated with ligation of the iron by the O atom are restricted to the active site, the rest of the molecule being unaffected. Binding of oxygen to the iron in compound II is associated with small but distinct changes in the porphyrin conformation. The angles between the *A*, *B* and *D* pyrrole rings and pyrrole *C* change by between 3 and 5° (Table 3). There is a significant change in the bridging pyrroles *B* and *C*. The aromatic rings Phe138 and Phe146 above the haem exhibit small movements associated with the ligand binding and the haem adjustments. The largest (more than 0.25 Å) movement is at Phe146, which appears to be associated with the movement of pyrrole *B*. The Phe138 ring has significantly higher thermal parameters than its surrounding residues in the room-temperature and in the vitrified native structures. The movement of this ring in compound II is barely significant.

In compound II the second water (W36) molecule present in the active site of MLC is shifted just enough to avoid repulsive van der Waals contact with the iron-bound O atom at a distance of 3.48 Å. The *B* value of this water (25 Å^2) is a little higher than that of the surrounding atoms (about 20 Å^2). Unlike MLC-0.88, there is no other positive electron density in the active site that could be described as a partially occupied water. The highest positive electron density in the active site of compound II is around the CHC atom of haem, which shows an even higher bend of the haem at this position.

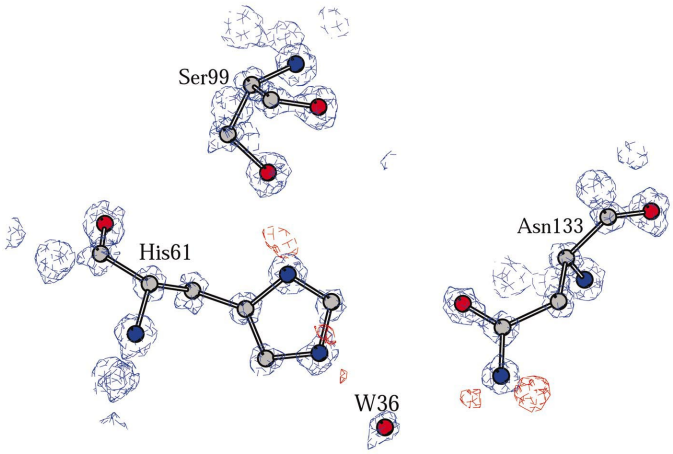

Figure 5
Experimentally observed H atoms on His61 and on Asn133 ND2. The $2F_o - F_c$ map contoured at 5σ is shown in blue and the $F_o - F_c$ map contoured at the 4σ level is shown in red. Both electron densities were calculated using MLC-0.88 data. The orientation of the active-site residues is confirmed as that surmised from previous X-ray data at lower resolution.

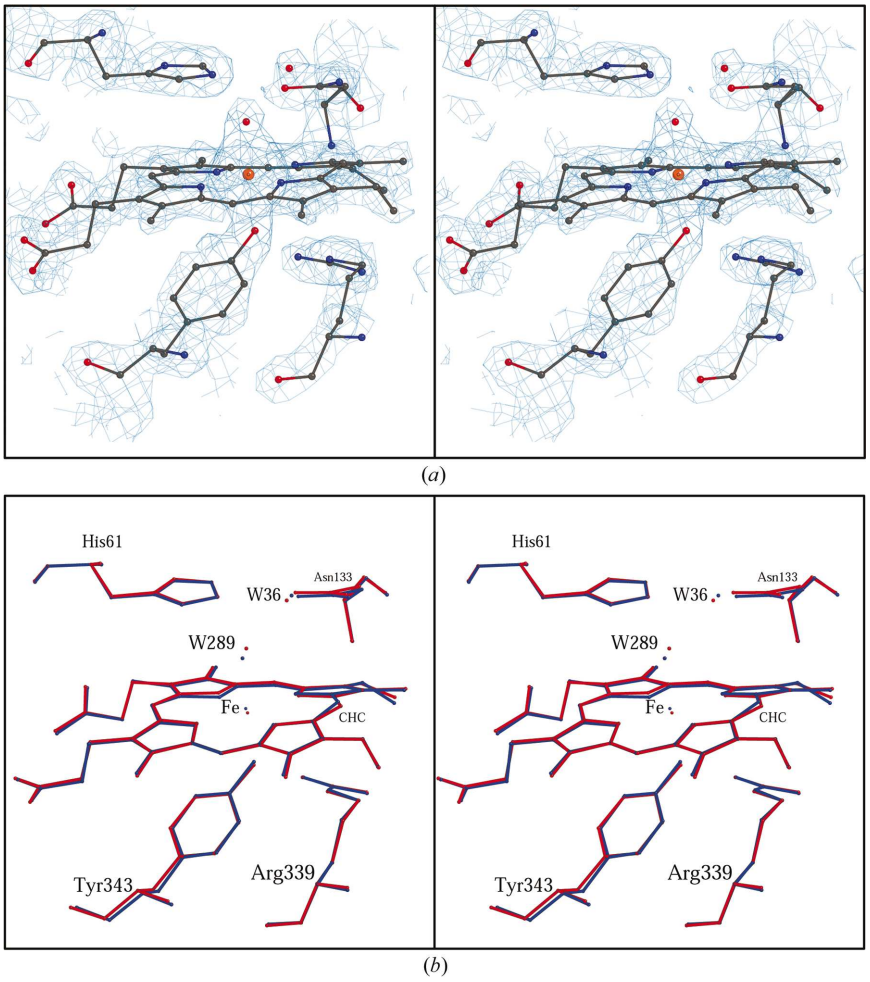

Figure 6
(*a*) Electron density demonstrating oxygen binding to the haem Fe atom (orange) in compound II. C atoms are coloured grey, O atoms red and N atoms blue. The final $2F_o - F_c$ map (contoured at 1σ level) is in blue. (*b*) Comparison of the active site in resting MLC-0.88 (red) with MLC–PAA (blue).

In compound II the Fe atom moves 0.24 Å toward the sixth ligand. Tyr343 OH moves in concert by 0.28 Å, accompanied by small changes of the dihedral angle at CA. The distance between the sixth liganded O atom and the iron is 1.87 Å (during refinement this distance was not restrained and van der Waals repulsion for this particular pair of atoms was removed from the restraint list) compared with the 2.2–2.3 Å distance to the water in the resting state (Table 2). In the compound I and II PMC structures at low resolution (Gouet *et al.*, 1996) this distance was restrained to be 1.71 Å. Inspection of the coordinates shows that in these structures there is distortion of the iron's octahedral coordination not found in MLC. The iron coordination seems to be near-ideal octahedral (Fig. 6). There is no significant change in the iron–phenolic O distance in the two enzyme states. There is a movement of the iron relative to the pyrrole nitrogen plane. In the resting state, the deviation of the iron from the pyrrole nitrogen plane is 0.14 Å towards the proximal side and in compound II it is 0.04 Å towards the distal side of the haem.

Below the haem, residue Arg339 adjusts in concert with the movements in the haem that accompany oxygen binding. Further away from the haem there appear to be few significant changes in the structure, which can be correlated with the creation of an electron-transport pathway.

3.2.7. The haem channel. As in other catalases, the MLC active site is connected with the surface of the molecule by a 25 Å long narrow channel. There are subtle but probably important differences between the channel in MLC and the other catalases. Instead of the wide entrance funnel in BLC (which becomes narrow at about 15 Å from the haem iron) most of the MLC funnel is filled by the C-terminal residues, which form a narrow entrance (Fig. 7). This narrow entrance (the distance between atoms forming it is about 6–7 Å) is composed of the hydrophobic residues Ala164, Ala453 and two residues from close to the C-terminus, Ile492 and Val495. The narrow entrance of the channel may increase discrimination between substrates of various sizes, in a good agreement with the evidence that MLC is slower for larger substrates and faster for small substrates relative to other catalases (Jones & Middlemiss, 1972).

The second narrow region of the channel is common to all known catalase structures. It is situated about 7 Å from the haem iron. This neck is probably responsible for discrimination and orientation of the substrate while it is entering the active site. As in other catalases, this part of the channel is formed by the hydrophobic residues Val101, Val114, Phe138 and Ile150, with distances between nearest wall atoms of 5.9–

8.5 Å. Just above the neck lie Gln153 and Gln167, each of which display two conformations (rotated by 180° around χ_3). Difference maps calculated after the refinement of one of the two conformations show positive and negative densities at the positions occupied by the OE1 and NE2 atoms in MLC-0.88. At lower resolution the MLC-RT comparison of *B* values of these atoms after refinement also clearly supports the possibility of two conformations. The *B* value of N becomes less and O becomes more than that of surrounding atoms. A comparison of *B* values of these atoms in compound II shows that this tendency remains unchanged. While this disordered behaviour may serve to accelerate diffusion in the haem channel of MLC, the equivalent residues to Gln153 and Gln167 are not conserved among catalases.

Comparison of *B* values with surrounding atoms suggests that only two water molecules at the extreme ends of the haem channel are nearly fully occupied, W36 in the active site and W292 in the channel near to the molecular surface. Unsurprisingly, they are conserved in all MLC structures (Fig. 7). W292 is situated at the beginning of the channel near to the surface of the molecule and makes strong hydrogen bonds with the main-chain carbonyl O atom of Arg495 and the amide of Pro497, which would help stabilization of the C-terminus and channel structure. After inclusion of these waters in the structure-factor calculations, the difference map exhibits positive density showing that, in spite of the high resolution, not all the contents of the channel can be taken to describe discrete atoms. Partially occupied waters and positive electron density may indicate a dynamic structure for this haem-linked water system. We believe that the positions of partially occupied water molecules in the active-site channel represent the path of substrate and product into and out of the active site. Indeed, there is experimental evidence (Melik-Adamyan *et al.*, 2001) showing that single-site mutation of the residue corresponding to MLC's His61 to Ala or Asn in HPII inactivates the enzyme and simultaneously increases the order of the water molecules in the active-site channel.

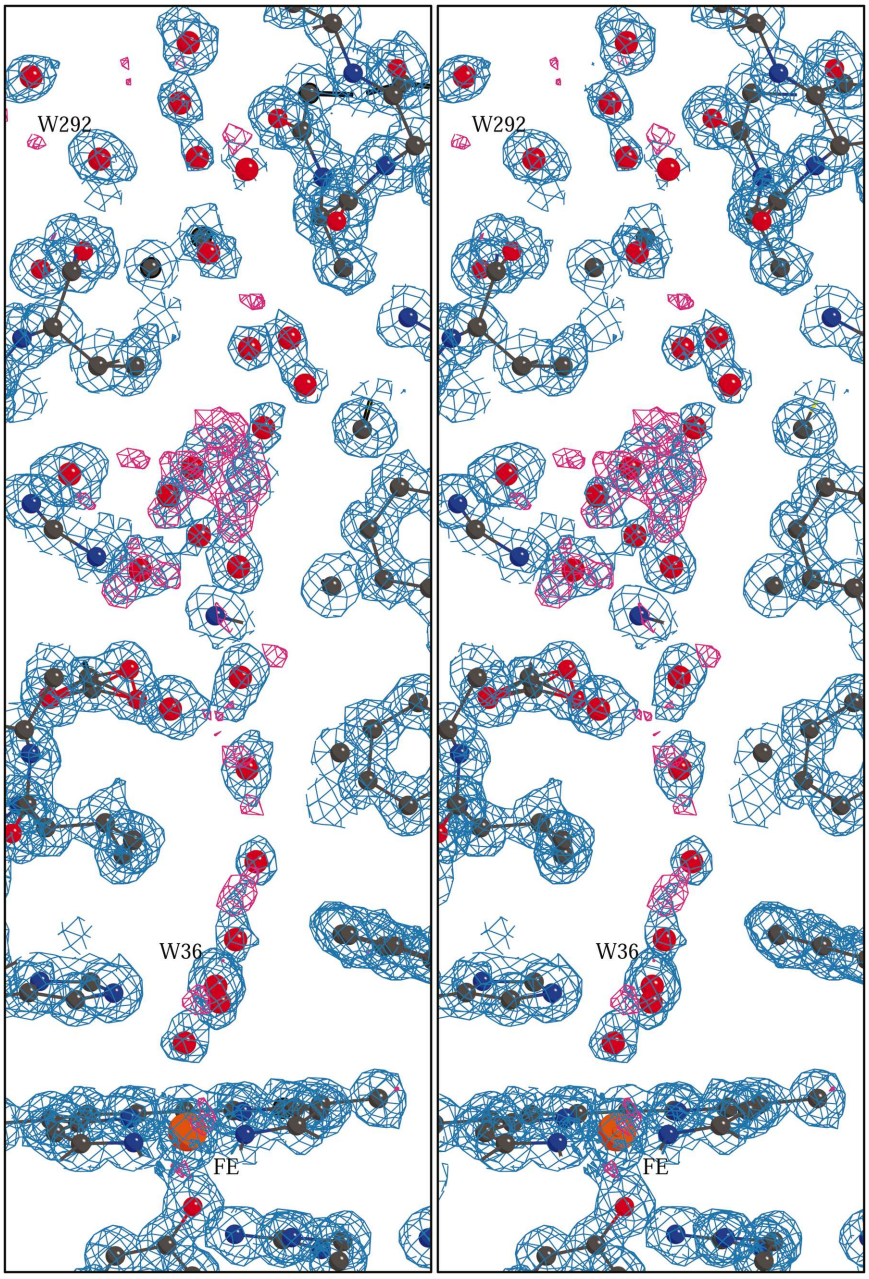

Figure 7
The electron density calculated using MLC-0.88 data in the haem channel and in the active site. The $2F_o - F_c$ map contoured at the 1σ level is shown in blue and the $F_o - F_c$ map contoured at 4σ is shown in red. Atom colours are as in Fig. 6. The water structure suggests easy access between active site and molecule surface for small substrates and reaction products. Inherent mobility in this channel is likely to be crucial for the enzyme to achieve the rapid reaction rates for which it is renowned.

Part of this work was performed during GNM's PhD studentship in the Institute of Crystallography, Moscow, Russia and a visit to EMBL, Hamburg, Germany supported by a fellowship from EMBL. We thank Dr S. V. Shlyapnikov for protein N-terminal sequencing. This work was supported by the BBSRC (Structural Biology Centre Grant 87/SB09829), by grant No. 99-02-16191 from the Russian Foundation for Basic Research and by the Royal Society joint project with the former Soviet Union (reference rc/jp/jun98). GNM is supported by a Wellcome Trust Senior Fellowship. We thank the European Commission for support of the work at EMBL, Hamburg through the HCMP Access to Large Installations Project.

References

Almarsson, O., Sinha, A., Gopinath, E. & Bruice, T. C. (1993). *J. Am. Chem. Soc.* **115**, 7093–7102.

Andersson, L. A., Johnson, A. K., Simms, M. D. & Willingham, T. R. (1995). *FEBS Lett.* **370**, 97–100.

Antonyuk, S. V., Melik-Adamyan, W. R., Popov, A. N., Lamzin, V. S., Hempstead, P. D., Harrison, P. M., Artymiuk, P. J. & Barynin, V. V. (2000). *Crystallogr. Rep.* **45**, 105–116.

Barynin, V. V., Whittaker, M. M., Antonyuk, S. V., Lamzin, V. S., Harrison, P. M., Artymiuk, P. J. & Whittaker, J. W. (2001). *Structure*, **9**, 725–738.

Bradford, B. U., Seed, C. B., Handler, J. A., Forman, D. T. & Thurman, R. G. (1993). *Arch. Biochem. Biophys.* **303**, 172–176.

Bravo, J., Fita, I., Gouet, P., Jouve, H. M., Melik-Adamyan, W. & Murshudov, G. N. (1997). *Oxidative Stress and the Molecular Biology of Antioxidant Defenses*, edited by G. Scanalious, pp. 407–446. New York: Cold Spring Harbor Laboratory Press.

Bravo, J., Verdaguer, N., Tormo, J., Betzel, C., Switala, J., Loewen, P. C. & Fita, I. (1995). *Structure*, **3**, 491–502.

Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.

Chance, B. & Oshino, N. (1971). *Biochem. J.* **122**, 225–233.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cruickshank, D. W. J. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 11–22. Warrington: Daresbury Laboratory.

Deisseroth, A. & Dounce, A. L. (1970). *Physiol. Rev.* **50**, 319–375.

DeLuca, D. C., Dennis, R. & Smith, W. G. (1995). *Arch. Biochem. Biophys.* **320**, 129–134.

Dolphin, D., Forman, A., Borg, D. C., Fajer, J. & Felton, R. H. (1971). *Proc. Natl Acad. Sci. USA*, **68**, 614–618.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Fita, I. & Rossmann, M. G. (1985). *J. Mol. Biol.* **185**, 21–37.

Fita, I., Silva, A. M., Murthy, M. R. N. & Rossmann, M. G. (1986). *Acta Cryst.* B**42**, 497–515.

Gouet, P., Jouve, H.-M. & Dideberg, O. (1995). *J. Mol. Biol.* **249**, 933–954.

Gouet, P., Jouve, H-M., Williams, P. A., Andersson, I., Andreoletti, P., Nussaume, L. & Hajdu, J. (1996). *Nature Struct. Biol.* **3**, 951–956.

Halliwell, B. & Gutteridge, J. M. C. (1984). *Biochem. J.* **219**, 1–14.

Hendrickson, W. A. & Konnert, J. H. (1980). *Biomolecular Structure, Function, Conformation and Evolution*, edited by R. Srinavasan, pp. 43–58. New York: Pergamon Press.

Herbert, D. & Pinsent, J. (1948). *Biochem. J.* **43**, 193–202.

Hillar, A., Nicholls, P., Switala, J. & Loewen, P. C. (1994). *Biochem. J.* **300**, 531–539.

Jones, P. & Middlemiss, D. N. (1972). *Biochem. J.* **130**, 411–415.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kirkman, H. N., Rolfo, M., Ferraris, A. M. & Gaetani, G. F. (1999). *J. Biol. Chem.* **274**, 13908–13914.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.

Lanir, A. & Schejter, A. (1975). *FEBS Lett.* **55**, 254–256.

Lardinois, O. M., Mestdagh, M. M. & Rouxhet, P. G. (1996). *Biochim. Biophys. Acta*, **1295**, 222–238.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.*, **26**.

Mallery, S. R., Bailer, R. T., Hohl, C. M., Ngbautista, C. L., Ness, G. M., Livingston, B. E., Hout, B. L., Stephens, R. E. & Brierley, G. P. (1995). *J. Cell. Biochem.* **59**, 317–328.

Marie, A. L., Parak, F. & Hoppe, W. (1979). *J. Mol. Biol.* **129**, 675–676.

Mate, M. J., Murshudov, G. N., Bravo, J., Melik-Adamyan, W. R., Loewen, P. C. & Fita, I. (2001). *Handbook of Metalloproteins*, edited by A. Messerschmidt, R. Huber & T. Poulos, pp. 486–502. Chichester: John Wiley & Sons.

Mate, M. J., Zamocky, M., Nykyri, L. M., Herzog, C., Alzari, P. M., Betzel, C., Koller, F. & Fita, I. (1999). *J. Mol. Biol.* **286**, 135–149.

Melik-Adamyan, W. R., Bravo, J., Carpena, X., Switala, J., Mate, M. J., Fita, I. & Loewen, P. C. (2001). *Proteins Struct. Funct. Genet.* **44**, 270–281.

Murshudov, G. N., Grebenko, A. I., Barynin, V. V., Dauter, Z., Wilson, K. S., Vainshtein, B. K., Melik-Adamyan, W. R., Bravo, J., Ferran, J. M., Ferrer, J. C., Switala, J., Loewen, P. C. & Fita, I. (1996). *J. Biol. Chem.* **271**, 8863–8868.

Murshudov, G. N., Melik-Adamyan, W. R., Grebenko, A. I., Barynin, V. V., Vagin, A. A., Vainshtein, B. K., Dauter, Z. & Wilson, K. S. (1992). *FEBS Lett.* **312**, 127–131.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.

Nicholls, P., Fita, I. & Loewen, P. C. (2001). *Adv. Inorg. Chem.* **51**, 51–106.

Ogata, M. (1991). *Hum. Genet.* **86**, 331–340.

Oldfield, T. J. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 67–74. Warrington: Daresbury Laboratory.

Olson, L. P. & Bruice, T. C. (1995). *Biochemistry*, **34**, 7335–7347.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Putnam, C. D., Arvai, A. S., Bourne, Y. & Tainer, J. A. (2000). *J. Mol. Biol.* **296**, 295–309.

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.

Roberts, J. E., Hoffman, B. M., Rutter, R. & Hager, L. P. (1981). *J. Biol. Chem.* **256**, 2118–2121.

Schonbaum, G. R. & Chance, B. (1976). *The Enzymes*, Vol. 13, edited by P. D. Boyer, pp. 363–408. New York: Academic Press.

Studier, F. W. (1991). *J. Mol. Biol.* **219**, 37–44.

Vagin, A. A., Murshudov, G. N. & Strokopytov, B. V. (1998). *J. Appl. Cryst.* **31**, 98–102.

Vainshtein, B. K., Melik-Adamyan, W. R., Barynin, V. V., Vagin, A. A., Grebenko, A. I., Borisov V. V., Bartels, K. S., Fita, I. & Rossmann, M. G. (1986). *J. Mol. Biol.* **188**, 49–61.

Vuillaume, M. (1987). *Mutat. Res.* **186**, 43–72.

Welinder, K. G. (1991). *Biochim. Biophys. Acta*, **1080**, 215–220.

Zamocky, M. & Koller, F. (1999). *Prog. Biophys. Mol. Biol.* **72**, 19–66.